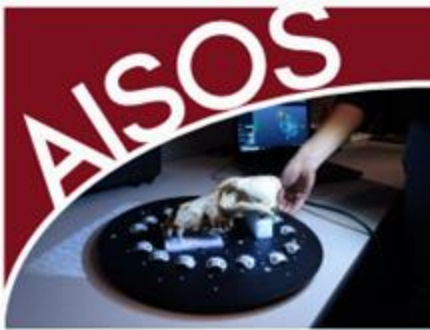


Data Sharing in the Humanities and Social Sciences

Alicia Hofelich Mohr, Ph.D.
College of Liberal Arts LATIS



How can AISOS help with your project?

AISOS (Advanced Imaging Service for Objects and Spaces) provides both methodological expertise and technologies for advanced imaging research.

We support

- Scanning Electron Microscopy (SEM)
- 3D Imaging Photogrammetry
- Reflected Light Capture
- White Light Profilometry
- Structured Light Imaging
- Gigapixel Imaging
- Terrestrial LiDAR
- Unity
- And More!

Software

Blender Agisoft Metashape Pro Dragonfly
Adobe Photoshop and Illustrator zBrush
MountainsMap Zereine Stacker
Geomagic Design X

have an idea? email aisos@umn.edu

SPECIALTIES AND SERVICES PROVIDED BY LATIS RESEARCH

Partnering with Latis Research from the beginning of any project ensures your team benefits from each of our specialties



STUDY PLANNING

Assist with Data Management Plans, IRB, infrastructure selection



DATA COLLECTION

Develop, program, and host survey, experimental, and mixed-method instruments. Assist with APIs, web scraping, and database development.



DATA ANALYSIS

Reshape, combine, and preprocess data from multiple sources. Assist in the use of qualitative, quantitative, or mixed methods tools.

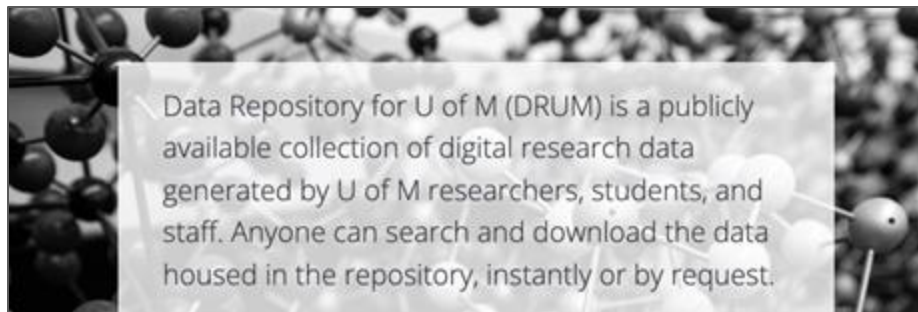


RESEARCH COMPUTING

Design computing solutions for grants, including a clustered compute environment, virtual machines, and support for computational code.

DATA CURATION NETWORK

Ethical. Reusable. Better.



Data Repository for U of M (DRUM) is a publicly available collection of digital research data generated by U of M researchers, students, and staff. Anyone can search and download the data housed in the repository, instantly or by request.

"Ideal" Scientific Data Sharing



Explore data

Search



Who we are

What we do

Join us

Help

Login

Sources of coaster brook trout (*Salvelinus fontinalis*) revealed by genomic analysis of brook trout populations along Minnesota's shoreline with Lake Superior

Mamoozadeh, Nadya ¹ ; Goldsworthy, Cory ²; Miller, Loren ²; Peterson, Nick ²; Meek, Mariah ¹

Author affiliations

Published Jul 03, 2023 on Dryad. <https://doi.org/10.5061/dryad.w9ghx3ftq>

Cite this dataset

Mamoozadeh, Nadya et al. (2023). Sources of coaster brook trout (*Salvelinus fontinalis*) revealed by genomic analysis of brook trout populations along Minnesota's shoreline with Lake Superior [Dataset]. Dryad. <https://doi.org/10.5061/dryad.w9ghx3ftq>

Data files	
Jul 03, 2023 version files	19.57 KB
id_highest_maf_snp.R	3.38 KB
README.txt	992 B
sample_meta...tchery.csv	2.74 KB
sample_meta...a.wild.csv	12.45 KB
Download full dataset	

Related works
Primary article
https://doi.org/10.10...6/j.jglr.2023.06.005

Script reproduces analysis

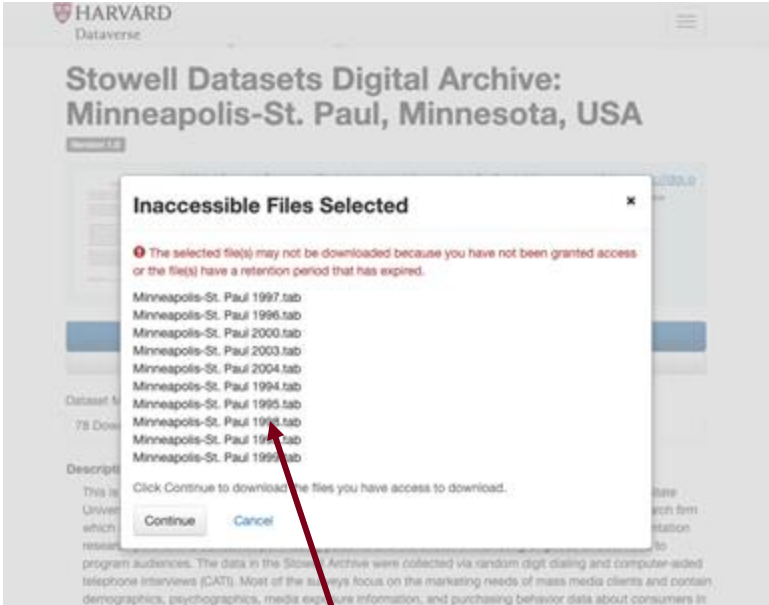
Data files + documentation capture all information

- ✓ Open access
- ✓ Standard metadata
- ✓ DOI

Social Science Challenges

QDR

The Qualitative Data Repository



Time/cost for researcher to grant access to restricted files

Individual deposit fees vary depending on the size and content of the data. Typical deposit fees range from \$300 to \$1,500. QDR calculates deposit fee estimates through the associated consultation with QDR staff. We are happy to discuss the estimated deposit fee, and any available fee waivers and discounts at your convenience. We connect with scholars who are planning on submitting funding proposals to help them understand the curation / preservation costs they should include in their proposal budget.

OPENICPSR

whether the data collection will be made available for public download or published as restricted data.

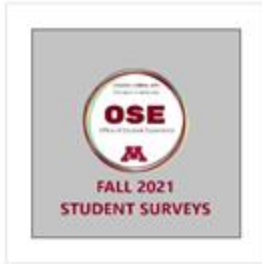
Users are required to apply for access to restricted data. Additionally, non-members will be charged \$550 to access the data, plus either \$349 for secure download or \$484 for virtual data enclave (VDE) administrative fees.

Fees to cover costs to repositories to vet access

Social Sciences Challenges

Fall 2021 University of Minnesota College of Liberal Arts First-Year Student Surveys

Estrella, Emma; Davis, Cassandra N; Mody, Isha; DeWitt, Katie; Roy, Aarushi; Yan, Chenwei; Hines, Alexander; Hammell, Abbey E; Hofelich Mohr, Alicia; Ronning, Emily
(2022-01-07)



Title

Fall 2021 University of Minnesota College of Liberal Arts First-Year Student Surveys

[View this record in Altmetric](#) [View this record in ORCID](#)

File View/Open	Description	Size
Readme.txt	Readme file	19.89 KB
F21_Student_Survey_1_questions.pdf	Survey 1 questions	112.38 KB
F21_OSE_Student_Survey_1_data.csv	Survey 1 data	358.22 KB
F21_Student_Survey_2_questions.pdf	Survey 2 questions	120.07 KB
F21_OSE_Student_Survey_2_data.csv	Survey 2 data	318.19 KB
F21_OSE_Student_Survey_recoding_codes.csv	Recoding information for demographic variables	15.71 KB

8. De-identification: These survey responses represent a sample of the CLA FYE population. The demographic variables were recorded (see included recode document) to reduce disclosure risk. Any "Prefer not to respond" response selections were recoded as missing. Local suppression was also for the demographic variables used to meet 2-anonymity for each datasets. This was done using the `sdcmicro_5.6.1` package in R 4.2.0 on macOS Big Sur 11.6.6. The `sdcmicro` suppression reports are below.

Survey 1:
The input dataset consists of 1180 rows and 22 variables.
--> Categorical key variables: `Demo_1_race_recoded`, `Demo_1_ethnicity_recoded`, `Demo_2_recoded`, `Demo_3`, `Demo_4_recoded`, `IS_FGEN`

Information on categorical key variables:

Infos on 2/3-Anonymity:

Number of observations violating
- 2-anonymity: 0 (0.000%) | in original data: 83 (7.034%)
- 3-anonymity: 30 (2.542%) | in original data: 127 (10.763%)
- 5-anonymity: 106 (8.983%) | in original data: 190 (16.102%)

Local suppression:

KeyVar	Suppressions (#)	Suppressions (%)
<code>Demo_1_race_recoded</code>	16	1.356
<code>Demo_1_ethnicity_recoded</code>	0	0.000
<code>Demo_2_recoded</code>	10	0.847
<code>Demo_3</code>	3	0.254
<code>Demo_4_recoded</code>	56	4.746
<code>IS_FGEN</code>	0	0.000

Survey 2:
The input dataset consists of 869 rows and 29 variables.
--> Categorical key variables: `Demo_1_race_recoded`, `Demo_1_ethnicity_recoded`, `Demo_2_recoded`, `Demo_3`, `Demo_4_recoded`, `IS_FGEN`

Recoding & de-identification effort to prepare for public sharing

Social Science Challenges



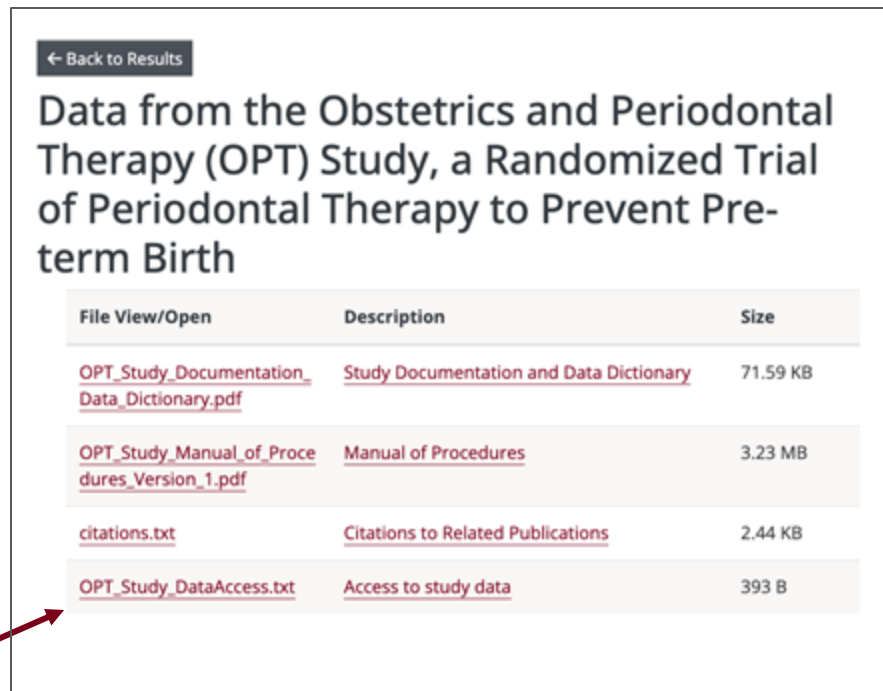
University Digital Conservancy [Communities](#) [Browse ▾](#) [Statistics](#) [About ▾](#) [Contact & Collections](#)

Data Repository for the University of Minnesota (DRUM)

DRUM Human Participant Data Policy

Data involving human participants will be subject to additional review to ensure that it meets our requirements for public sharing. Data that we determine is insufficiently de-identified or which has not properly consented participants to the sharing of individual-level data in a public repository may be rejected from DRUM, or may be subject to removal after acceptance.

Effort to ensure older files meet current ethics

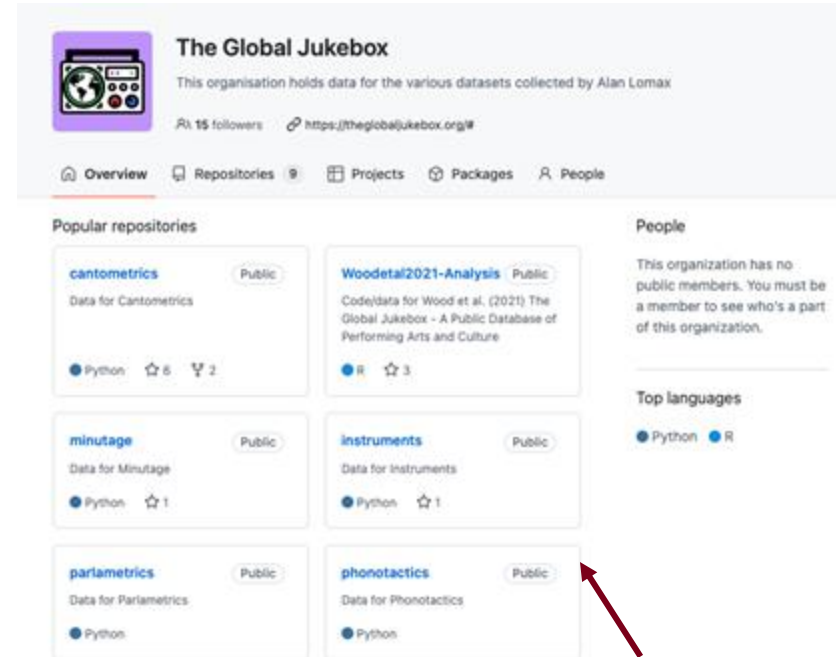
A screenshot of a DRUM record page. At the top left is a button labeled '← Back to Results'. The main title is 'Data from the Obstetrics and Periodontal Therapy (OPT) Study, a Randomized Trial of Periodontal Therapy to Prevent Pre-term Birth'. Below the title is a table with three columns: 'File View/Open', 'Description', and 'Size'. The table contains four rows of data. A red arrow points from the text 'Effort to ensure older files meet current ethics' to the 'OPT_Study_DataAccess.txt' row in the table.

File View/Open	Description	Size
OPT_Study_Documentation_Data_Dictionary.pdf	Study Documentation and Data Dictionary	71.59 KB
OPT_Study_Manual_of_Procedures_Version_1.pdf	Manual of Procedures	3.23 MB
citations.txt	Citations to Related Publications	2.44 KB
OPT_Study_DataAccess.txt	Access to study data	393 B

Humanities Challenges



Immersive and interactive web experience



Very different from "repository" version

Humanities Challenges

Shared via websites
for interactivity



Minnesota Historical Election Archive
Curated by **smart** POLITICS

Explore over 160 years
of election results in
Minnesota.

20,461 Elections
15,927 Candidates
17 Offices

Search for a candidate or office (optional) Search

Show: Elections Candidates
From: Enter a year

American Classical
Hymns
AN ANTHOLOGY OF PRE-CIVIL WAR HYMN TUNES BASED
ON CLASSICAL MUSIC

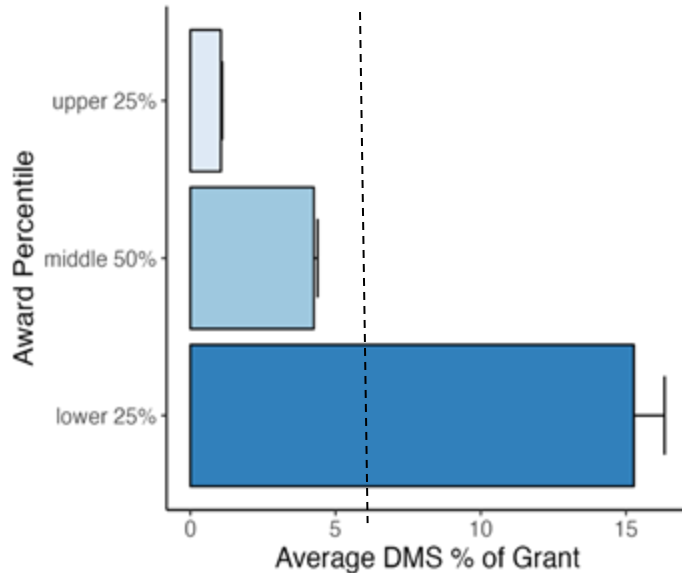
HOME EXPLORE THE ANTHOLOGY TUNES IN CHORAL PERFORMANCE

ABOUT THE PROJECT ABOUT US

American Classical Hymns

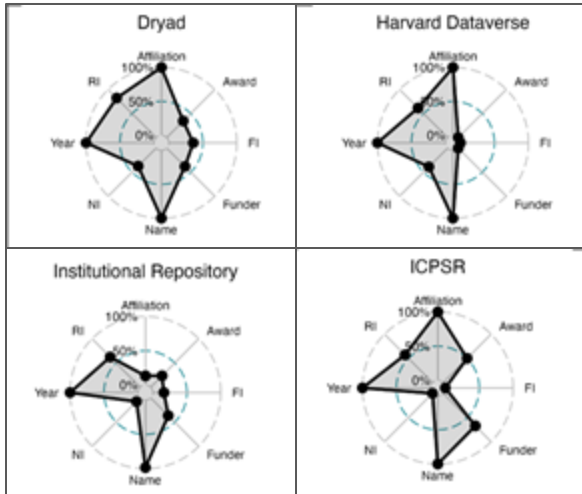
THE ANTHOLOGY: CHORAL SHOWCASE

Costs of data management & sharing for small grants



- Realities of Academic Data Sharing (RADS) survey/interview of 255 PIs with NIH, NSF, and DOE grants
- Participants divided by award size into smaller (lower 25%) to larger (upper 25%) grants
- Overall average cost of DMS was 6%, but near 15% for smaller awards

It's hard to find and keep track of it all



- Pulled data citations from DataCite and CrossRef DOIs for six institutions
- 90% of data sets were associated with 10 data publishers
- Inconsistent and missing metadata (plus no DOIs) lead to missing many others
- Top 10 publishers also lacked OSTP encouraged metadata fields

Summary



<https://www.storyblocks.com/images/stock/image-of-pink-piggy-bank-surrounded-by-stacks-of-dollar-bills-sqgliudquwj6gt34br>

- The cost of managing and sharing data is a relatively large proportion of grant budgets for small awards
- Social sciences and humanities face additional, long term costs for data sharing
 - Granting access requests for restricted sharing
 - Maintaining security and standards for old datasets
 - Preparing data for sharing
 - Web/technical support and hosting beyond grant
- Global infrastructures designed to track and catalog data publishing may miss social sciences and humanities data, especially for those shared in non-repository format

Thank you!

Contact: hofelich@umn.edu